Structure
Ways & Means

# On the Relationship between Diffraction Patterns and Motions in Macromolecular Crystals

Peter B. Moore[1,*]
[1]Department of Chemistry, Yale University, New Haven, CT 06520-8107, USA
*Correspondence: peter.moore@yale.edu
DOI 10.1016/j.str.2009.08.015

## SUMMARY

**The quality of many macromolecular crystal structures published recently has been enhanced through the use of new methods for treating the effects of molecular motion and disorder on diffraction patterns, among them a technique called translation, libration, screw-axis (TLS) parameterization. TLS parameterization rationalizes those effects in terms of domain-scale, rigid-body motions and, interestingly, the models for molecular motion that emerge when macromolecular diffraction data are analyzed this way often make sense biochemically. Here it is pointed out that all such models should be treated with caution until it is shown that they are consistent with the diffuse scatter produced by the crystals that provided the diffraction data from which they derive.**

## INTRODUCTION

In macromolecular crystal structures all atoms are assigned exact locations and crystal structures are generally interpreted assuming those positions are immutable. In fact, the positions of atoms in macromolecular crystals are constantly changing due to thermal motions and the PDB files in which the atomic coordinates of macromolecular crystal structures are recorded include information about those motions. It is contained in the lists of B factors (also known as ''Debye-Waller factors'' or ''temperature factors''), which are proportional to the variances of atomic positions in crystals. Atoms with big B factors are likely to be found far from their average positions, which are the positions reported for them in PDB files. Atoms with small B factors stay closer to home.

Atomic motions reduce the intensities of the reflections in diffraction patterns because they make the instantaneous structures of crystals vary from one unit cell to the next. The more accurately the effects of these motions are accounted for during refinement, the more accurate the structures that emerge. However, the larger the amplitudes of these motions, the harder it is to do that accounting. The reason is that motions depress the intensities of Bragg reflections and the effect increases with scattering angle. Thus the larger the amplitudes of the atomic motions in a crystal, the lower the resolution of the data that can be collected from it, and the lower that resolution, the smaller the number of parameters that can be extracted from the data. For the average macromolecular crystal structure, data limitations normally allow atomic motions to be characterized using only a single parameter, a B factor, which is proportional to the variance of an atom's excursions away from its mean position, averaged over all directions. Only when the resolution of the data is exceptionally high is it normally possible to take account of the fact that the motions of atoms in crystals are anisotropic.

The connection between resolution and structure refinement just described explains the increasing popularity of the translation, libration, screw-axis (TLS) approach (Cruickshank, 1956; Schomaker and Trueblood, 1968) for treating disorder in macromolecular crystals. TLS parameterization enables crystallographers to allow for the anisotropies of atomic motions when solving the structures of crystals that diffract to resolutions much lower than would otherwise permit, and the quality of structures often improves significantly when they do so (e.g., see, Kuriyan and Weiss, 1991; Winn et al., 2001).

Used simply as a tool for optimizing structures, TLS parameterization is, at worst, innocuous. If the free R factors of a structure falls when it is refined by TLS methods, it must be that the anisotropic B factors obtained better describe the crystal's diffraction characteristics than the isotropic B factors that they replace. However, the premise of TLS parameterization is that most of the atomic motions in crystals result from small, random rotations, i.e., ''librations,'' and translations of assemblies of atoms that, to first approximation, behave mechanically as rigid bodies. Thus when applied to macromolecular crystals, TLS parameterization yields descriptions of exactly the kinds of domain-scale motions that interest biochemists and, not surprisingly, biochemical significance has sometimes been imputed to them (e.g., Holbrook and Kim, 1984; Holbrook et al., 1985; Howlin et al., 1989; Papiz et al., 2003; Chaudhry et al., 2004; Korostelev and Noller, 2007).

The purpose of this essay is to remind the reader that the observation that the free R factor of some crystal structure went down when it was refined using some TLS-derived model for molecular dynamics does not prove the model is correct, although it may be as good a model as the data allow. The existence of other, quite different models that are as good or better cannot be excluded a priori. That said, it is comparatively easy to find out whether a particular dynamics model is plausible. This can be done by comparing the diffuse scatter a crystal produces with the diffuse scatter that model predicts. Before functional significance is ascribed to any such model, it would seem reasonable to ask that its proponents demonstrate its plausibility by this means or, alternatively, make a few frames of data available to the public so that those who are interested can make such comparisons.

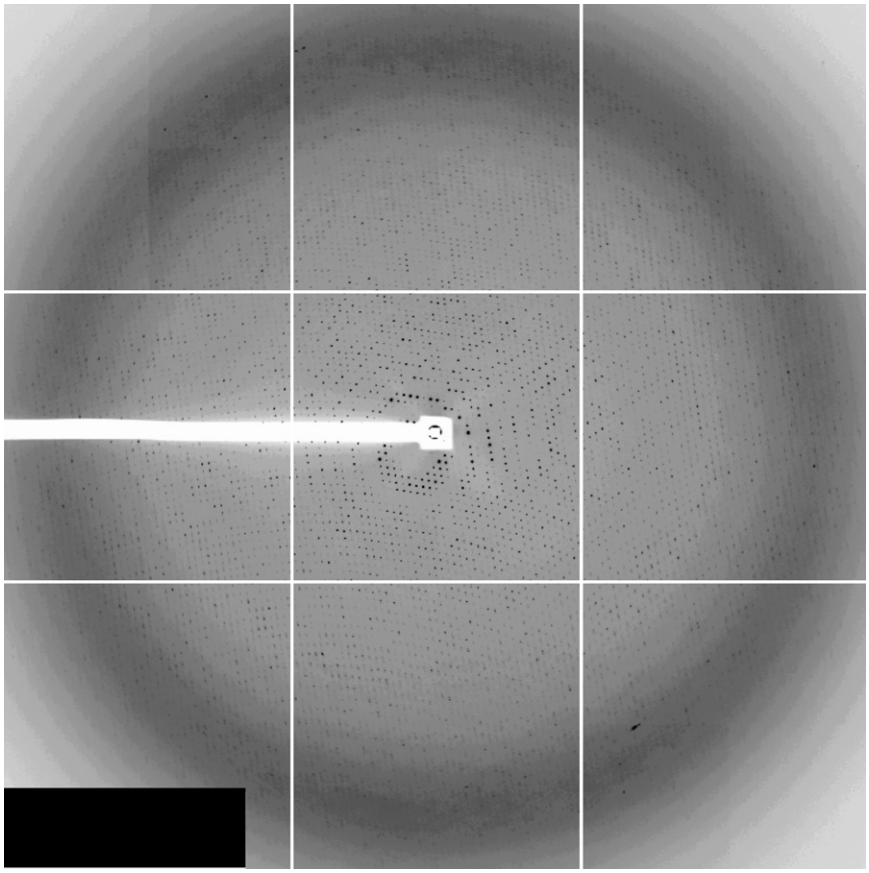

**Figure 1. A typical Frame of Macromolecular Diffraction Data**
This frame of data is part of a data set collected from a crystal of 70S ribosomes from *Thermus thermophilus*. The frame shown is a 0.3° oscillation image obtained at APS on 24-ID-C, using a 3×3 CCD detector. (The division of the image into nine squares is obvious.) The wavelength of the X-rays used was 0.9795Å, the detector was 500 mm from the crystal, and the resolution at the (horizontal) edge of the image is 0.3125 $Å^{-1}$. This image was provided by R. Evans and G. Blaha.

## RESULTS

### Diffuse Scattering Patterns and Molecular Motion

It has been known for roughly a century that unit cell-to-unit cell variation in the positions of corresponding atoms, whether the result of thermal motion or static disorder, makes crystals diffract X-rays in directions other than those consistent with von Laue's equations (James, 1965). The component of this non-Bragg scatter relevant here is called "diffuse scatter" or "thermal diffuse scatter," and it is responsible for the quasi-continuous background on which the Bragg reflections in diffraction patterns are superimposed. Figure 1, which is discussed at greater length below, is a single frame of oscillation data obtained from a crystal of 70S ribosomes from *Thermus thermophilus*. The thermal diffuse scatter in this frame is dominated by a broad ring of intensity at a Bragg spacing of about 3.3 Å, and in this regard this frame is similar to many another. (Diffuse rings of this sort are described as "solvent rings" or "water rings" even though most of the scatter they represent is produced by crystal disorder rather than by solvent scatter per se [Glover et al., 1991].)

Unlike B factors, which report only on the variances of the average position assigned individual atoms, diffuse scattering patterns are sensitive to the physics responsible for those variances, and the effects are not subtle. Suppose crystals were obtained of some multiprotein complex that includes a lysozyme molecule and suppose, improbably, that lysozyme motions were the sole source of the diffuse scatter produced by those crystals. Figure 2A shows what the diffuse scatter would look like in a still image similar to Figure 1 if the lysozyme molecules in question were undergoing small, random, rigid-body translations of the sort postulated in TLS parameterizations. Figure 2B shows what the diffuse scatter would look like if the atoms in those lysozyme molecules were undergoing random translations of exactly the same magnitude, but were doing so in a completely uncorrelated manner. The correct B factor to assign to all lysozyme atoms during the refinement of the structure of this hypothetical crystal would be exactly the same in both cases: 20 $Å^2$.

In the 1990s a lot of work was done on the diffuse scattering patterns produced by macromolecular crystals, which are an obvious feature of the diffraction data collected from such crystals at synchrotron light sources. It was shown that the character of the diffuse scattering these crystals produce varies significantly from one macromolecular crystal to the next; diffuse scattering patterns are not always as amorphous as the one displayed in Figure 1 (Glover et al., 1991). Furthermore, substantial advances were made in understanding the different kinds of disorder that might account for them (e.g., Doucet and Benoit, 1987; Clarage et al., 1992; Kolatkar et al., 1994; Thune and Badger, 1995; Wall et al., 1997; Clarage and Phillips, 1997; Héry et al., 1998; see also Benoit and Doucet, 1995; Moss and Harris, 1995). Nevertheless, surprisingly, there is still only a single paper that describes a study in which a TLS-derived model for macromolecular disorder was tested by comparing the diffuse scattering pattern it implies to real data (Perez et al., 1996).

### On the Relationship between Crystal Disorder and Diffuse Scatter

The theory of diffuse scattering can be approached several different ways (Doucet and Benoit, 1987; Clarage et al., 1992; Welberry and Butler, 1994; Mizuguchi et al., 1994). The simplest is the one described by James (1965), and it leads to expressions similar to those presented earlier by Mizuguchi et al. (1994), among others.

Omitting constants that are of no concern here, I(**S**,t), the intensity of the X-radiation scattered by some crystal at some location in reciprocal space, **S**, at time t, is:

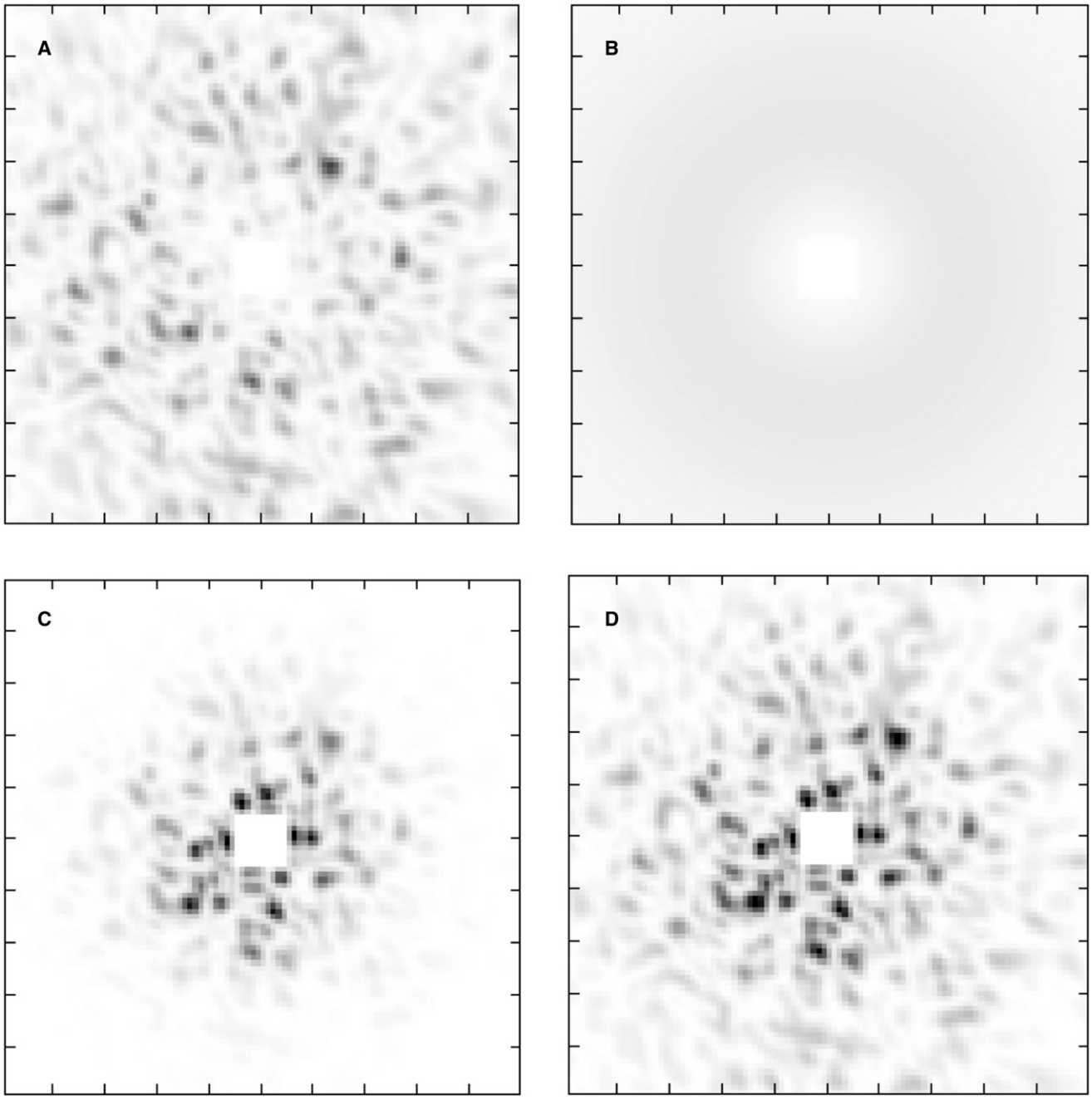


**Figure 2. The Effects of Translational Disorder on Diffuse Scattering**

In both this figure and Figure 3, the structure responsible for the diffuse scatter shown is hen egg white lysozyme and the coordinates used to represent that molecule are those reported for its tetragonal form by Nanao et al. (2005) (PDB 2BLX). All images show what the diffuse scatter would look like in a single still taken from the overall, three-dimensional scattering pattern of the protein. In every case, the X-ray beam is directed down the z axis of the protein and the wavelength is 1.0 Å. At the corner of each still $|(\mathbf{S})| = 0.707$ Å$^{-1}$. No allowance is made for solvent contrast effects.

(A) The diffuse scatter anticipated if the entire molecule moves as a rigid body and the motion can be described using an isotropic Gaussian, the variance of which is 0.25 Å$^2$ (B factor = 20 Å$^2$).

(B) The diffuse scatter expected if each atom in the molecule moves in exactly the same way as in (A) but does so independently.

(C) The average transform of lysozyme in crystals where lysozyme molecules are experiencing either the disorder that leads to the diffuse scatter shown in (A) or the disorder responsible for the diffuse scatter displayed in (B).

(D) The transform of lysozyme in the absence of any disorder whatever. To make these images easier to look at, their gray scales were set to different values. In (C) and (D), black implies $\geq$150,000 arbitrary units, while in (A), black is $\geq$100,000 units and in (B) it is $\geq$50,000 units. The pattern in (B) would be invisible otherwise.

$$I(\mathbf{S},t) = \sum\sum \exp(-2\pi i(\mathbf{u}_c - \mathbf{u}_{c'})\cdot\mathbf{S}) \times \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \exp(-2\pi i(\delta_{ic}(t) - \delta_{ic'}(t))\cdot\mathbf{S}) \quad (1)$$

The first two summations are over all unit cells, while $\mathbf{u}_c$ is the vector from the origin of the coordinate system to the origin of unit cell c and $\mathbf{u}_{c'}$ is the corresponding vector for unit cell c′. The second pair of summations both run over all atoms in the unit cell. $f_i$ and $f_j$ are atomic scattering factors of atoms i and j, the dependence of which on $|\mathbf{S}|$ we ignore here. ($|\mathbf{S}| = 2\sin\theta/\lambda$, where $\theta$ is half the scattering angle.) The average positions of atoms i and j in the unit cell are $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively. $\delta_{ic}(t)$ is the displacement of atom i from its average position in unit cell c at time t and $\delta_{jc'}(t)$ is the corresponding vector for atom j in unit cell c′. For every atom in every unit cell, $\delta_{ic}(t)$ times the probability of $\delta_{ic}(t)$, integrated over all unit cells and over all time, i.e., over its "displacement distribution function," is zero. This is also true of the average value of the component of $\delta_{ic}(t)$ in any given direction.

The Bragg component of the scattering of the crystal in question is:

$$I(\mathbf{S})_{Bragg} = \sum\sum \exp(-2\pi i(\mathbf{u}_c - \mathbf{u}_{c'})\cdot\mathbf{S}) \times \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \langle\exp(-2\pi i(\delta_{ic}\cdot\mathbf{S}))\rangle\langle\exp(-2\pi i(\delta_{ic'}\cdot\mathbf{S}))\rangle.$$

Expressions of the form $\langle \exp(-2\pi i((\delta_i \cdot \mathbf{S})) \rangle$ are evaluated by expanding the exponential as a power series and averaging, term by term, over the appropriate displacement distribution function. By definition, the average value of the first order term in $(\delta_i \cdot \mathbf{S})$ has to be zero, and hence the first two terms in the expansion are always:

$$\langle\exp(-2\pi i(\delta_i\cdot\mathbf{S}))\rangle = 1 - 2\pi^2\langle(\delta_i\cdot\mathbf{S})^2\rangle + \ldots\ldots,$$

where $\langle (\delta_i \cdot \mathbf{S})^n \rangle$ is the displacement distribution function-weighted average of the dot product of **S** and the displacement of atom i, raised to the nth power. Thus the first two non-zero terms in this power series are always the same as the first two terms in the power series expansion of $\exp(-2\pi^2\langle(\delta_i \cdot \mathbf{S})^2\rangle)$, and if the displacement distribution function of atom i is Gaussian, then:

$$\exp(-2\pi i\langle(\delta_i\cdot\mathbf{S})\rangle) = \exp\left(-2\pi^2\langle(\delta_i\cdot\mathbf{S})^2\rangle\right),$$

exactly. Thus it is usually reasonable to write:

$$I(\mathbf{S})_{Bragg} \approx \sum\sum \exp(-2\pi i(\mathbf{u}_c - \mathbf{u}_{c'})\cdot\mathbf{S}) \times \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \exp\left(-2\pi^2\left(\langle(\delta_{ic}\cdot\mathbf{S})^2\rangle + \langle(\delta_{jc'}\cdot\mathbf{S})^2\rangle\right)\right). \quad (2)$$

$$\approx T^2 \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \exp\left(-2\pi^2\left(\langle(\delta_i\cdot\mathbf{S})^2\rangle + \langle(\delta_j\cdot\mathbf{S})^2\rangle\right)\right),$$

where T is the total number of unit cells in the crystal. (The second version of Equation 2 is valid only when **S** obeys von Laue's equations.) The exponential decay terms in Equation 2 are conventional B factor expressions.

Equation 2 expresses the fact that Bragg scattering is determined by the average positions of atoms in unit cells because at Bragg refection positions in reciprocal space, the Fourier transform of the structure of a crystal is the linear sum of the Fourier transforms of the electron density distributions of all its unit cells, which must be T times the Fourier transform of the average electron density distribution. Thus at Bragg reflection positions in reciprocal space, $I(\mathbf{S})_{Bragg} = T^2(F_{ave}(\mathbf{S}))^2$. On average, atomic displacements are accounted for in real space by replacing atomic electron density distributions with the convolution of those electron density distributions with displacement distribution functions.

The scattering described by Equation 1 that is not accounted for by Equation 2, time and space averaged, is the diffuse scatter of a crystal, i.e., $I(\mathbf{S})_{dif} = \langle I(\mathbf{S},t) - I(\mathbf{S})_{Bragg} \rangle$. Hence:

$$I(\mathbf{S})_{dif} = \sum\sum \exp(-2\pi i(\mathbf{u}_c - \mathbf{u}_{c'})\cdot\mathbf{S}) \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \left[\langle\exp(-2\pi i(\delta_{ic}(t) - \delta_{jc'}(t))\cdot\mathbf{S})\rangle - \exp\left(-2\pi^2\left(\langle(\delta_{ic}\cdot\mathbf{S})^2\rangle + \langle(\delta_{jc'}\cdot\mathbf{S})^2\rangle\right)\right)\right]. \quad (3)$$

Once the first term in the square brackets in Equation 3 is evaluated, it is found that Equation 3 can be rewritten as follows:

$$I(\mathbf{S})_{dif} = \sum\sum \exp(-2\pi i(\mathbf{u}_c - \mathbf{u}_{c'})\cdot\mathbf{S}) \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \exp\left(-2\pi^2\left(\langle(\delta_{ic}\cdot\mathbf{S})^2\rangle + \langle(\delta_{jc'}\cdot\mathbf{S})^2\rangle\right)\right) \times \left[\exp\left(4\pi^2\langle(\delta_{ic}(t)\cdot\mathbf{S})(\delta_{jc'}(t)\cdot\mathbf{S})\rangle\right) - 1\right]. \quad (4)$$

Thus if the displacements of two (different) atoms do not correlate, their motions will not contribute to diffuse scatter because the expression in square brackets will be 0.

To better understand the implications of Equations 3 and 4, consider two extreme cases. First, suppose that in some crystal the displacements of different atoms do not correlate at all. In that case, the only terms in the double sum that will not be zero are those that relate atoms to themselves and the diffuse scatter will obey the following equation:

$$I(\mathbf{S})_{dif} = T\sum f_i^2\left[1 - \exp\left(-4\pi^2\langle(\delta_i\cdot\mathbf{S})^2\rangle\right], \quad (5)$$

a well-known result (James, 1965). However, if the motions of atoms in the different unit cells of some crystal are uncorrelated, but within unit cells some group of atoms is translating thermally as a rigid body, then:

$$I(\mathbf{S})_{dif} = T\left[1 - \exp\left(-4\pi^2\langle(\delta_{i'}\cdot\mathbf{S})^2\rangle\right)\right] \times \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}), \quad (6)$$

where the two sums run over all atoms in the rigid assembly in question and $\delta$ is the displacement vector for the entire rigid body (Perez et al., 1996). (If there are several such rigid bodies in the unit cell, each undergoing independent motion, the diffuse scattering pattern produced by the crystal will be the sum of the diffuse scattering patterns produced by each such unit separately.) Figures 2B and 2A illustrate the difference between Equations 5 and 6, respectively.

If the disorder in some crystal is purely translational then in reciprocal space the average value of $I(\mathbf{S})_{dif}$ in any shell of constant $|\mathbf{S}|$ will be the same no matter whether atoms are translating independently or translating as rigid groups. Statistically, the difference between the resulting diffuse scattering patterns will lie in the variance of $I(\mathbf{S})_{dif}$ within shells, which will be zero if atoms are translating independently and $< I(\mathbf{S})_{dif} >$ ($< I(\mathbf{S})_{dif} > \sim\sum f_i^2$) if atoms are translating as rigid assemblies, as can easily be shown using Wilson statistics. Thus, uncorrelated motion produces smooth diffuse scattering distributions while correlated motion results in "lumpy" diffuse scattering distributions.

Even though intensity distributions like the one in Figure 2A look random, they have a length scale and it is manifest in the distance between adjacent maxima, which is roughly the same everywhere in reciprocal space. That reciprocal space distance is approximately the reciprocal of half the linear dimensions of the objects undergoing correlated motion. In this instance, the inter-peak distance corresponds to a distance of ~20 Å, which is about the radius of lysozyme.

As noted earlier, in TLS parameterization procedures macromolecules are treated as assemblies of rigid groups of atoms that undergo small, random translations, rotations, and screw axis motions relative to each other. In addition, it is assumed that there are no correlations between the motions occurring in different unit cells and that each rigid group moves independently (Winn et al., 2003). Thus, if the disorder in some crystal really is TLS disorder, the diffuse scatter the crystal produces should be describable using a more general form of Equation 6:

$$I(\mathbf{S})_{dif} = T\sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \left[\exp\left(-2\pi^2 <(\delta_i(t) - \delta(t))\cdot\mathbf{S})^2>\right) - \exp\left(-2\pi^2\left(<(\delta_i\cdot\mathbf{S})^2> + <(\delta_j\cdot\mathbf{S})^2>\right)\right)\right]. \quad (7)$$

(Equation 7 is obtained from Equation 3 by assuming [1] that molecular motions correlate only within unit cells, [2] that the displacement distribution functions of corresponding atoms are the same in all unit cells, and [3] that $< \exp(-2\pi i\mathbf{x}) > = \exp(-2\pi^2 < \mathbf{x}^2 >)$.)

## On the Diffuse Scatter Produced by Rotational Disorder

If the diffuse scatter observed in some crystal results from rigid-body librations of some assembly about a single axis only, then to first order in $< \varepsilon^2 >$, the variance of the distribution of rotations, Equation 7 will have the following form:

$$I(\mathbf{S})_{dif} = T\sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j)\cdot\mathbf{S}) \times \left[\exp\left(-2\pi^2((\mathbf{r}_i - \mathbf{r}_j)\mathbf{x}\mathbf{1}\cdot\mathbf{S})^2 <\varepsilon^2>\right) - \exp\left(-2\pi^2\left(\left((\mathbf{r}_i\mathbf{x}\mathbf{1}\cdot\mathbf{S})^2 + <(\mathbf{r}_j\mathbf{x}\mathbf{1}\cdot\mathbf{S})^2\right)<\varepsilon^2>\right)\right)\right]. \quad (8)$$

In this equation, $\mathbf{r}_i$ and $\mathbf{r}_j$ are vectors from an origin on the axis of rotation to atoms i and j. The vector **1** is a unit vector coincident with the axis of the rotation. If more than one axis of libration has to be taken into account, the first and second terms in square brackets must be replaced by products of terms of similar form, one for each rotation axis, and if the assembly is also undergoing random, rigid-body translations, then the second term in the square brackets of Equation 8 must additionally be multiplied by $\exp(-4\pi^2(< (\delta\cdot\mathbf{S})^2 >))$. If the motion of concern is a screw motion, the first term in square brackets will be the same as in Equation 8, i.e., it will have the same form it would if the motion was a pure rotation. However, for both atoms considered, the second term in square brackets will be:

$$\exp\left(-2\pi^2\left((\mathbf{r}_i\mathbf{x}\mathbf{1}\cdot\mathbf{S})^2\right) + 2a(\mathbf{r}_i\mathbf{x}\mathbf{1}\cdot\mathbf{S})(\mathbf{1}\cdot\mathbf{S}) + \mathbf{a}^2(\mathbf{1}\cdot\mathbf{S})^2\right)<\varepsilon^2>\right)$$

where a is the advance in the direction of the axis of rotation per radian of rotation.

Greater insight into the effects of librational disorder can be obtained if Equation 8 is rewritten in cylindrical polar coordinates with the z axis coincident with the axis of rotation of the rigid group of concern. In cylindrical polar coordinates, the Fourier transform of a molecule is:

$$F(R,\Psi,Z) = \sum \exp(in(\Psi - \pi/2)) \times \sum f_i J_n(2\pi R r_i)\exp(in\varphi_i)\exp\left(-2\pi i Z z_i\right)$$

(Cochran et al., 1952; Klug et al., 1958). For objects of arbitrary structure, the first sum, which is a sum over Bessel function orders, runs from $n = -\infty$ to $n = +\infty$. $J_n$ is the Bessel function of order n and $f_i$ is the structure factor of the ith atom, which is located at $(r_i, \varphi_i, z_i)$. The second sum runs over all atoms. If a molecule is subject to rigid-body librations, then the Fourier-Bessel equivalent of Equation 8 will be:

$$I(R,\Psi,Z)_{dif} = T\sum\sum \exp(i(n-m)(\Psi - \pi/2)) \times \sum\sum f_i f_j J_n(2\pi R r_i) J_m(2\pi R r_j)\exp\left(i(n\varphi_i - m\varphi_j)\right) \times \exp(-2\pi i Z(z_i - z_j)) \times \left[\exp\left(-(1/2)(n-m)^2<\varepsilon^2>\right) - \exp\left(-(1/2)(n^2+m^2)<\varepsilon^2>\right)\right].$$

By making a substitution that is standard in the Fourier-Bessel field (i.e.,

$$g_n(R,Z) = \sum f_i J_n(2\pi R r_i)\exp(in\varphi_i)\exp(-2\pi i Z z_i),)$$

a simpler equation can be obtained:

$$I(R,\Psi,Z)_{dif} = T\sum\sum \exp(i(n-m)(\Psi - \pi/2)) \times \left[\exp\left(-(1/2)(n-m)^2<\varepsilon^2>\right) - \exp\left(-(1/2)(n^2+m^2)<\varepsilon^2>\right)\right] g_n(R,Z) g_m^*(R,Z). \quad (9)$$

It is easy to show that in cylindrical polar coordinates, the Bragg component of the total scatter is:

$$I(R,\Psi,Z)_{Bragg} = T^2\sum\sum \exp(i(n-m)(\Psi - \pi/2)) \times \left[\exp\left(-(1/2)(n^2+m^2)<\varepsilon^2>\right)\right] g_n(R,Z) g_m^*(R,Z). \quad (10)$$

Comparison of Equation 9 and Equation 6 shows that the diffuse scatter produced by librational disorder is qualitatively different from the diffuse scatter generated by translational displacements. Within any shell of constant $|\mathbf{S}|$, the distribution of intensities in diffuse scattering patterns caused by rigid-body translations is

the same as that in the transform of the object experiencing those translations, as Figure 2 illustrates. Figure 2D is a single still from the transform of a lysozyme molecule, assuming no disorder whatever. Allowing for a small difference in gray scales (see the figure legend), Figure 2D (= $\sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{S})$; see Equation 6) is clearly the sum of Figure 2A, the diffuse scatter that would be observed in that same still if the protein was subject to that same amount of translational disorder (= $[1 - \exp(-4\pi^2(<(\delta \cdot \mathbf{S})^2>))] \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{S})$), and Figure 2C, the transform of that same molecule attenuated by that same amount of translational disorder (= $+ \exp(-4\pi^2(<(\delta \cdot \mathbf{S})^2>)) \sum\sum f_i f_j \exp(-2\pi i(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{S})$).

In contrast, it is impossible to recognize the parent transform in the diffuse scattering patterns caused by rigid-body libration, e.g., compare Figures 2D and 3B . The reason is that librational diffuse scatter lacks all contributions from the $J_o$ component of the transform of the object that is librating, which makes sense physically because the $J_o$ term in Equation 9 is the Fourier-Bessel transform of the cylindrically averaged structure of that object. By definition, that aspect of its structure is invariant to rotation around z.

Figure 3 presents some diffuse scattering patterns that further explore the impact of rigid-body librations on diffuse scatter. The object undergoing rigid-body rotations is a lysozyme molecule, as before, and in all panels in Figure 3 its average orientation with respect to the incident X-ray beam is the same and the same as it is in all panels of Figure 2. Figure 3A shows the diffuse scatter predicted if the molecule is translating normal to an axis parallel to the incident X-ray beam that passes through its center of gravity, while librating around that same axis. $<\varepsilon^2>$ is $(2^\circ)^2$, while $<|\delta|^2>$ is 0.25 Å$^2$. This panel is to be compared with Figure 2A, which it obviously resembles. (The disorder responsible for the diffraction pattern in Figure 2A is translational disorder only.) Figure 3B is plotted on a gray scale that has 1/5 the range of the gray scale in Figure 2A, and the diffraction pattern it displays is caused by librations of the same amplitude and orientation that contributed to pattern displayed in Figure 3A. Figure 3C is just like Figure 3B except that the axis of libration has been displaced perpendicular to the X-ray beam by 5 Å. Figure 3C faintly resembles Figure 2A, presumably because the rotation responsible for it translates the center of gravity of the protein to some degree. Figure 3D shows the diffuse scattering predicted if the rotation axis is perpendicular to the incident X-ray beam instead of parallel to it. The orientation of the rotation axis is clearly horizontal in this instance.

### On the Validation of Models for Molecular Dynamics

What should be done to validate TLS models or other models for the dynamics of molecules in crystals experimentally? In the first place, if no biochemical conclusions are to be drawn from them, validation is unnecessary, for the reasons given in the Introduction. However, when that is not the case, models of this sort should be held to a higher standard. For example some unprocessed frames of data might be made available to the public, perhaps as part of the supplemental material for a publication, so that readers can judge whether the diffuse scatter manifest in those frames is qualitatively consistent with there being domain-scale, rigid body motions in the crystals in question.

Although obviously desirable, it would be much more challenging to test models for molecular motions by comparing predicted diffuse scattering patterns with measured data quantitatively. Methods for collecting the data required have been worked out. For example, Wall et al. (1997) have shown how to measure the nonspherically symmetric part of the diffuse scatter produced by protein crystals. That component of the overall diffuse scatter is the component of greatest interest here because it should be non-zero only if the atoms in a crystal are engaged in correlated motions.

Of much greater concern is the computation of the diffuse scattering patterns to compare with measured data. Until algorithms are discovered that are far more efficient than those I have devised (see below), the computation of complete diffuse scattering patterns will be practical only for crystals of small proteins, e.g., lysozyme. Run on a computer that has a 2.4 MHz, quad-core cpu and 16 Mbytes of memory, the programs I have written take ~15 min to compute a single frame of diffuse data for a domain the size of lysozyme undergoing TLS motions and for a domain the size of the 30S ribosomal subunit it takes ~36 hr to compute 20% of a single such frame. These difficulties not withstanding, I would feel that the credibility of a TLS-derived model for the motions of domains in a large, multidomain structure had been strongly enhanced if its proponents were to show that it predicts the nonspherical part of the diffuse scatter of even a modest portion of a single frame of real oscillation data.

## DISCUSSION

As has often been pointed out in the past, and as Equation 2 confirms, the effects of disorder on the intensities of the reflections in a crystal's diffraction pattern are entirely explained by the displacement parameters of individual atoms taken one at a time. Thus it is impossible to infer anything about the correlations that may or may not exist between the motions of atoms in a crystal if the only data taken into account are the intensities of Bragg reflections. All models for those motions that predict similar temperature factors will be equivalent. Figures 2A and 2B illustrate the implications of that critical point.

TLS parameterization is generally embarked on after the structure of some crystal has been solved and partially refined and at which point all of its atoms will have been assigned isotropic B factors (Winn et al., 2003). Guided by chemical intuition, the user then defines the domains in the structure that are to be treated as rigid bodies during TLS parameterization. Alternatively, programs now available on the TLSMD server (http://skuld.bmsc.washington.edu/~tlsmd/) can be used to decide how best to divide molecules into domains. However it is done, the set of parameters describing the motions of those domains that best accounts for the observed isotropic B factors in some least-squares sense is then found. If the temperature factors of the atoms in some domain are all similar, that fact will be interpreted as evidence for rigid-body translations. If the temperature factors of atoms on the surface of the domain are larger than those of the atoms in its interior, which they are liable to be in any case (Brooks et al., 1988), that will be taken as evidence of librations. Thus no matter what is actually going on in a molecule, a set of TLS parameters is guaranteed to emerge. Before additional rounds of refinement are embarked upon, those parameters are used to compute anisotropic thermal

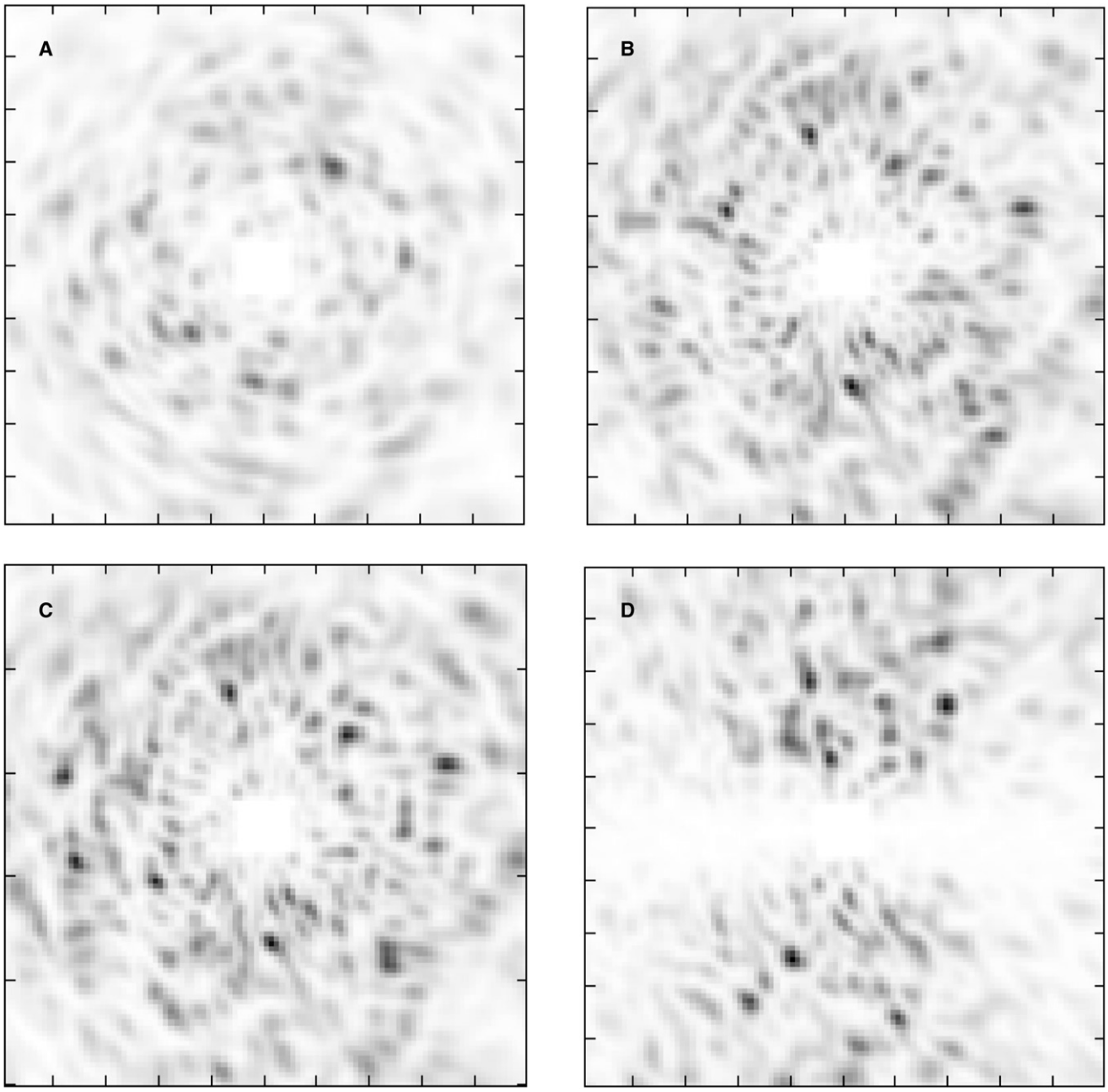


**Figure 3. The Effects of Rotational Disorder on Diffuse Scattering Patterns**

See the legend of Figure 1 for details.

(A) The diffuse scatter shown results from a combination of isotropic, rigid-body, translational disorder that has a B factor of 20 Å$^2$ (see Figure 1A) and a Gaussian rotational disorder that has a variance of $(4^\circ)^2$. The librations in question are about an axis parallel to the incident X-ray beam. Black is $\geq$100,000 units.

(B) The diffuse scatter produced by rotational component of the disorder that contributes to the diffuse scatter shown in (A). Black is $\geq$20,000 units.

(C) This image displays the effect of rotational disorder of the same magnitude as that in (B) except that while the direction of the axis of rotation is the same as in (B), it passes through a position in the protein that is 5 Å from the protein's center of gravity.

(D) In this instance, the axis of rotation passes through the center of gravity of the protein, but is perpendicular to the X-ray beam rather than being parallel to it. The magnitude of the rotation is the same as in all other panels.

ellipsoids for all atoms and once that is done, the means by which those ellipsoids were obtained are of no consequence.

The reason TLS refinement can improve structures is not that the models for domain motions used are necessarily correct, although they may be, but rather that the motions of atoms in most macromolecular crystals are anisotropic, and TLS parameterization provides a way of accounting for that anisotropy, at least approximately. Good things happen when this is done; electron density maps become easier to interpret and free R factors fall. Since the number of parameters that need to be

determined during TLS parameterization is generally orders of magnitude smaller than the number that would otherwise have to be determined to assign triaxial thermal ellipsoids to all of the atoms in a structure individually, it can be applied to low resolution structures.

Obviously, appropriate identification of the rigid units in a structure is the key to obtaining TLS parameters that are not only crystallographically useful but also biochemically relevant. For small molecule crystallographers this is seldom a problem; chemical logic can be relied on. In a small number of instances, macromolecular crystallographers have been able to identify rigid units the same way, with happy results (Holbrook and Kim, 1984; Holbrook et al., 1985). However, it is much more difficult to be sure what parts of a large macromolecular structure will behave mechanically as rigid bodies, if any. But suppose that a TLS model is obtained that works at both the crystallographic and the biochemical level. Can one then safely conclude that the model in question is a usefully appropriate description of the dynamical behavior of that macromolecule in its crystals?

It is not obvious that the answer to the question just raised is always going to be affirmative. For example, a recently reported TLS refinement of the structure for the 70S ribosome from *Thermus thermophilus* indicates that both ribosomal subunits librate significantly around axes that are roughly normal to the subunit interface (Korostelev and Noller, 2007). Thus one anticipates that the diffuse scatter produced by those crystals should resemble Figure 3B. (That expectation was confirmed by computing a small portion of the diffuse scattering pattern that would be seen if a rigid 30S ribosomal subunit in such a crystal were undergoing libration [data not shown].) However, in the diffuse scatter shown in Figure 1, which is a frame of data obtained from crystals of the same ribosomes, there is no indication whatever of large-scale, correlated motions. This discrepancy does not prove that the TLS model just referred to is incorrect because it derives from data obtained from tetragonal 70S crystals (I422), while Figure 1 shows data from orthorhombic crystals ($P2_12_12_1$) (G. Blaha, personal communication). However there are grounds for concern.

Clearly the acid test for this or any other proposal for the way the atoms move in a crystal is to see if it correctly predicts the diffuse scattering pattern obtained from that crystal. This conclusion is just as applicable to normal mode approaches for describing the motions of atoms in crystals as it is to TLS parameterization. Normal mode analysis is an alternative to TLS parameterization for obtaining anisotropic thermal ellipsoids for macromolecular crystal structures that have been solved to modest resolutions. In normal mode approach, computations are done on a crystal structure to identify its lowest energy, normal modes of vibration. Anisotropic temperature factors can be estimated for individual atoms using the displacements predicted for them assuming that those modes account for most of their motions (Kidera et al., 1994; Poon et al., 2007). The advantages of the normal mode method are that its users need not identify the rigid domains in macromolecules, if any, and that the molecular motions that emerge from it are guaranteed to make chemical sense, in detail, which TLS motions need not, even when the domains in some molecule have been appropriately identified.

The inverse problem, namely, the deduction of models for molecular motions from diffuse scattering patterns, is difficult, as experience with lysozyme crystals has amply demonstrated. Unlike the diffuse scattering shown in Figure 1, the diffuse scattering patterns produced by lysozyme crystals are highly structured and it follows that atomic motions in these crystals are correlated. Caspar and colleagues, who were pioneers in this field, proposed a simple model for such motions, namely that the motions of pairs of atoms in a protein are highly and positively correlated if they are neighbors and that the degree of correlation falls off exponentially with the distance between them (Clarage et al., 1992). Using that model, they computed diffuse scattering profiles for lysozyme that resemble the ones produced by both its triclinic and tetragonal crystals (Clarage et al., 1992). A few years later, Benoit and colleagues addressed the diffuse scattering from the same tetragonal crystals using a TLS approach and, not surprisingly, the diffuse scattering profiles they computed resemble Figure 3A (Perez et al., 1996). (Their paper includes an interesting critique of the exponential decay model of Clarage et al. [1992]) Later on, Smith and coworkers analyzed the same system using molecular dynamics and concluded that a good match between observation and computation could be obtained by treating $\leq\sim5$ residue sequences within lysozyme as dynamic clusters (Héry et al., 1998). Their computed diffuse scattering patterns appear to fit the data better than those of their predecessors, but the fit was still imperfect. Thus three different dynamical models have been found that could plausibly explain the diffuse scatter produced by lysozyme crystals, at least qualitatively. There is certainly still work to be done on this system, and it may be that, in lysozyme, atomic motions correlate in ways that defy simple description.

## Computations

In the course of these investigations, several simple-minded Fortran programs were written to compute diffuse scattering patterns. They were used both to confirm the equations provide above and to produce Figures 2 and 3. In the initial stages of the work, planes in reciprocal space were computed using two different programs, one based on Equation 8 and the other on Equation 9. The two programs gave results that agreed numerically to (±) 1%, which, given the several numerical approximations used in both programs, may correspond to round-off error. It should be noted that while in theory the sum over Bessel function orders in Equation 8 should run from $n = -\infty$ to $n = +\infty$, usefully accurate results will emerge if a finite range of orders is considered, e.g., $\sim-50$ to $\sim+50$, because the contributions of high order terms is negligible at modest values of $|\mathbf{S}|$. Even so, the run times for the Fourier-Bessel program were much longer than those of the Fourier-Sine program. (To make computations run faster, it was assumed that the form factors of all atoms have the same shape as that of carbon.)

Figures 2 and 3 show what the diffuse scattering intensity distributions would look like in single stills taken from the three-dimensional data sets obtained from a hypothetical protein one subunit of which is a single molecule of lysozyme. Implicitly the space group is P1. (Computations done by Benoit and colleagues [Perez et al., 1996] show that diffuse scattering patterns are just as lumpy as the ones shown in Figures 2 and 3 when there are several macromolecular domains moving

independently in the unit cells of a crystal, not just one. This observation is important because it suggests that the TLS scatter profiles obtained from multidomain macromolecules is unlikely to be smoothed by averaging.) Qualitatively, these still images give the same impression as the sections that were computed. In a few instances, frames of oscillation data were computed, rather than stills. While oscillation does smooth diffuse scattering patterns to some extent, the effect is too small to make oscillation frames look qualitative different from stills.

Given how slowly the Fourier-Bessel program ran, it made no sense to produce a version that would compute stills and so Figures 2 and 3 are based on Equation 9. Those interested in computing diffuse scattering profiles for their own crystals may find it simpler to compute $I(\mathbf{S})_{dif}$ by using the fast Fourier transform routines found in modern macromolecular crystallography packages to calculate $< I(\mathbf{S},t) - I(\mathbf{S})_{Bragg} >$ directly, as recommended by Benoit and coworkers (Perez et al., 1996). It should be noted that for geometrical reasons, the intensity of diffuse scattering patterns fall off systematically with scattering angle as $\cos^3(2\theta)$. Figures 2 and 3 were generated using GNUPLOT (http://www.gnuplot.info/docs/gnuplot.html).

(The programs referred to above will be made available upon request, even though they are far from polished and, from a user's point of view, extremely primitive. They may be useful to those interested in developing the computational tools that further investigations of diffuse scatter will so clearly require.)

## ACKNOWLEDGMENTS

I thank Robin Evans and Gregor Blaha for providing Figure 1 and Jimin Wang for his useful discussions. The computations required for this study were done in the Richards Center at Yale University with the help of its staff and with support provided by a grant from the National Institutes of Health (GM-022778).

Received: June 19, 2009
Revised: August 26, 2009
Accepted: August 28, 2009
Published: October 13, 2009

## REFERENCES

Benoit, J.P., and Doucet, J. (1995). Diffuse scattering in protein crystallography. Q. Rev. Biophys. *28*, 131–169.

Brooks, C., Karplus, M., and Pettitt, B.M., eds. (1988). Proteins: a theoretical perspective of dynamics, structure, and thermodynamics, *Volume 71* (New York: John Wiley).

Chaudhry, C., Horwich, A.L., Brunger, A.T., and Adams, P.D. (2004). Exploring the structural dynamics of the *E. coli* chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. J. Mol. Biol. *342*, 229–245.

Clarage, J.B., Larage, M.S., Phillips, W.C., Sweet, R.M., and Caspar, D.L.D. (1992). Correlations of atomic motions in lysozyme crystals. Proteins *12*, 145–157.

Clarage, J.B., and Phillips, G.N., Jr. (1997). Analysis of diffuse scattering and relation to molecular motion. Methods Enzymol. *277*, 407–432.

Cochran, W., Crick, F.H.C., and Vand, V. (1952). The structure of synthetic polypeptides I. The transform of atoms on a helix. Acta Crystallogr. *5*, 581–586.

Cruickshank, D.W.J. (1956). The analysis of the aniostropic thermal motion of molecules in crystals. Acta Crystallogr. *9*, 754–756.

Doucet, J., and Benoit, J.P. (1987). Molecular dynamics studied by the analysis of X-ray diffuse scattering from lysozyme crystals. Nature *325*, 643–646.

Glover, I.D., Harris, G.W., Helliwell, J.R., and Moss, D.S. (1991). The variety of diffuse scattering from macromolecular crystals and its respective components. Acta Crystallogr. B *47*, 960–968.

Héry, S., Genest, D., and Smith, J.C. (1998). X-ray diffuse scattering and rigid-body motion in crystalline lysozyme probed by molecular dynamics simulation. J. Mol. Biol. *279*, 303–319.

Holbrook, S.R., and Kim, S.-H. (1984). Local mobility of nucleic acids as determined from crystallographic data. I. RNA and B form DNA. J. Mol. Biol. *173*, 361–388.

Holbrook, S.R., Dickerson, R.E., and Kim, S.-H. (1985). Anisotropic thermal-parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method. Acta Crystallogr. B *41*, 255–262.

Howlin, B., Moss, D.S., and Harris, G.W. (1989). Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. Acta Crystallogr. A *45*, 851–861.

James, R.W. (1965). The Optical Principles of the Diffraction of X-Rays (Ithaca, NY: Cornell University Press).

Kidera, A., Matsushima, M., and Go, N. (1994). Dynamic structure of human lysozyme derived from X-ray crystallography: normal mode refinement. Biophys. Chem. *50*, 25–31.

Klug, A., Crick, F.H.C., and Wyckoff, H.W. (1958). Diffraction by helical structures. Acta Crystallogr. *11*, 199–213.

Kolatkar, A.R., Clarage, J.B., and Phillips, G.N., Jr. (1994). Analysis of diffuse scattering from yeast initator tRNA crystals. Acta Crystallogr. D Biol. Crystallogr. *50*, 210–218.

Korostelev, A., and Noller, H.F. (2007). Analysis of structural dynamics in the ribosome by TLS crystallographic refinement. J. Mol. Biol. *373*, 1058–1070.

Kuriyan, J., and Weiss, W.I. (1991). Rigid body motion as a model for crystallographic temperature factors. Proc. Natl. Acad. Sci. USA *88*, 2773–2777.

Mizuguchi, K., Kidera, A., and Go, N. (1994). Collective motion in proteins investigated by X-ray diffuse scattering. Proteins *18*, 34–48.

Moss, D.S., and Harris, G.W. (1995). Diffuse X-ray scattering from macromolecular crystals using synchrotron radiation. Radiat. Phys. Chem. *45*, 523–535.

Nanao, M.H., Sheldrick, G.M., and Ravelli, R.B. (2005). Improving radiation-damage substructure for RIP. Acta Crystallogr. D Biol. Crystallogr. *61*, 1227–1237.

Papiz, M.Z., Prince, S.M., Howard, T., Cogdell, R.J., and Isaacs, N.W. (2003). The structure and thermal motion of the B800-850 LH2 complex from *Rsp. acidophila* at 2.0 A resolution and 100 K: new structural features and functionally relevant motions. J. Mol. Biol. *326*, 1523–1538.

Perez, J., Faure, P., and Benoit, J.-P. (1996). Molecular rigid-body displacements in tetragonal lysozyme confirmed by X-ray diffuse scattering. Acta Crystallogr. D Biol. Crystallogr. *52*, 722–729.

Poon, B.K., Chen, X., Lu, M., Vyas, N.K., Quiocho, F.A., Wang, Q., and Ma, J. (2007). Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42 A crystallographic resolution. Proc. Natl. Acad. Sci. USA *104*, 7869–7874.

Schomaker, V., and Trueblood, K.N. (1968). On the rigid-body motion of molecules in crystals. Acta Crystallogr. B *24*, 63–76.

Thune, T., and Badger, J. (1995). Thermal diffuse x-ray scattering and its contribution to understanding protein dynamics. Prog. Biophys. Mol. Biol. *63*, 251–276.

Wall, M.E., Ealick, S.E., and Gruner, S.M. (1997). Three-dimensional diffuse x-ray scatter from crystals of *Staphylococcus* nuclease. Proc. Natl. Acad. Sci. USA *94*, 6180–6184.

Welberry, T.R., and Butler, B.D. (1994). Interpretation of difffuse X-ray scattering via models of disorder. J. Appl. Crystallogr. *27*, 205–231.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. *57*, 122–133.

Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. *374*, 300–321.

  